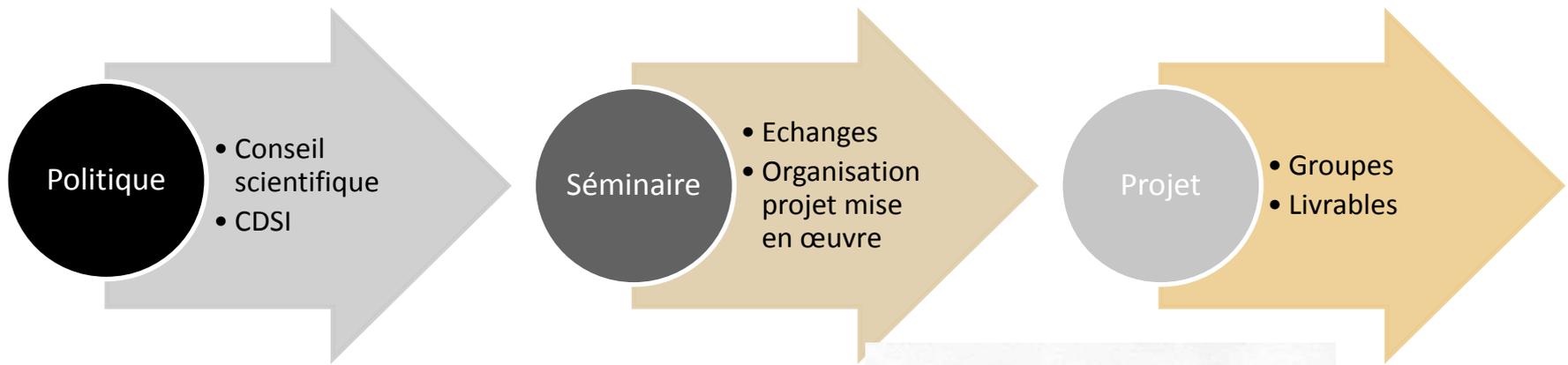


Partage des données de la recherche

Séminaire de lancement de la mise en œuvre de la politique Inra



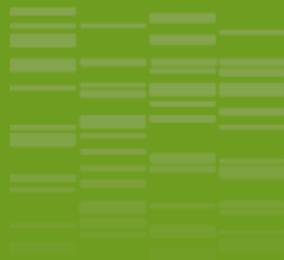
Le séminaire : lancement du projet de mise en œuvre de la politique



Plan



- ❖ Historique à l'Inra
- ❖ Contexte international
- ❖ Objectifs du séminaire
- ❖ Programme du séminaire



_01

Historique

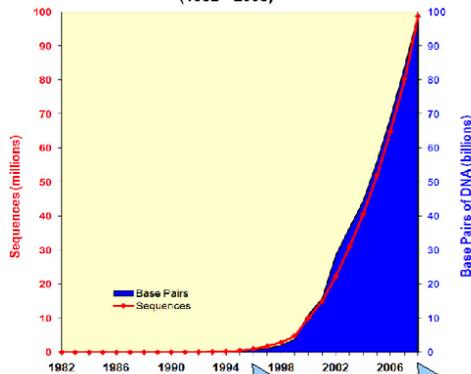
Quelques faits marquants à l'Inra

Sept. 2009: le CS s'interroge sur la question « Gestion et partage des données »

- ❖ Dans le domaine de la biologie/bioinformatique
- ❖ Dans un contexte de production massive des



Growth of GenBank (1982 - 2008)



606 seq
680,338 nt

1,021,211 seq
651,972,984 nt

98,868,465 seq
99,116,431,942 nt

Roche 454	Illumina	Solid
600 Mb	1300 Mb	3000 Mb
10 h	4 j	5j
400 pb / read	75 pb / read	35 pb / read
1982-1996	1982-1997	1982-1999

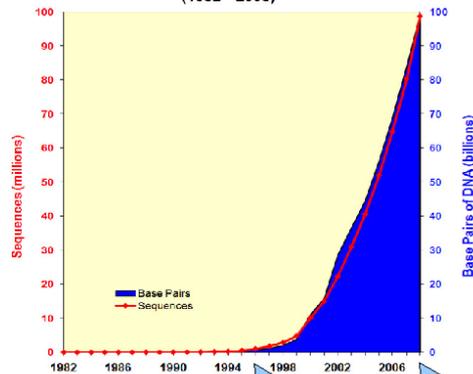
1 Run

Sept. 2009: le CS s'interroge sur la question « Gestion et partage des données »



- ❖ Dans le domaine de la biologie/bioinformatique
- ❖ Dans un contexte de production massive des données

Growth of GenBank (1982 - 2008)

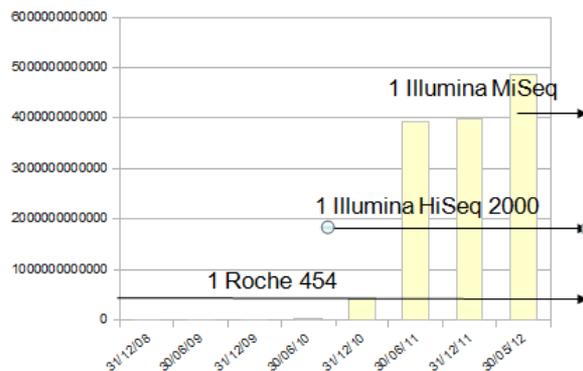


606 seq
680,338 nt

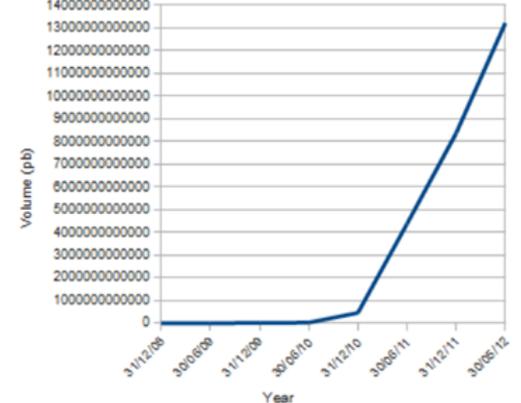
1,021,211 seq
651,972,984 nt

98,868,465 seq
99,116,431,942 nt

Evolution of volumes of data stored in NG6 every 6 months from the end of 2008 to May 2012



Evolution of total volume of sequencing data stored in NG6 from the end of 2008 to May 2012



2011-2012: le CS instruit la question « Gestion et partage des données »



❖ Groupe de travail piloté par D. Pontier

❖ Objectifs

- ✓ Regard scientifique sur les spécificités thématiques en matière d'acquisition de données, de besoins en matière de gestion et de partage à l'Inra
- ✓ Regard sur les changements induits par les mouvements du Big Data et de l'Open Data
- ✓ Analyse de quelques cadres nationaux et internationaux existants

2011-2012: le CS instruit la question « Gestion et partage des données »

❖ Conclusions (juin 2012): 9 recommandations

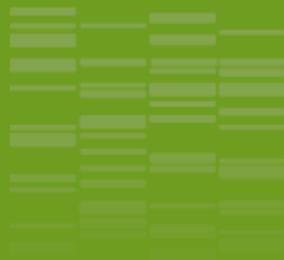
- ✓ R1 - Définir la politique de l'établissement et la communiquer
- ✓ R2 - Mettre en place un dispositif d'évaluation des données produites par l'Inra
- ✓ R3 - S'impliquer dans les comités internationaux de standardisation
- ✓ R4 - Développer un portail d'accès à un ensemble de ressources distribuées
- ✓ R5 - Prendre en compte le cycle de vie des données dès l'élaboration des projets de recherche

2011-2012: le CS instruit la question « Gestion et partage des données »

- ❖ Conclusions (juin 2012): 9 recommandations
 - ✓ R6 - Définir un cahier des charges pour les plateformes
 - ✓ R7 - Doter l'Inra d'infrastructures dimensionnées pour les stockages et calculs hautes performances
 - ✓ R8 - S'engager dans une politique de gestion des compétences répondant aux besoins en émergence
 - ✓ R9 - Conduire une réflexion inter-organismes
- ❖ Validation des recommandations par le CS en mai 2012

Sept-Dec 2012 : instruction en CDSI de la recommandation R1

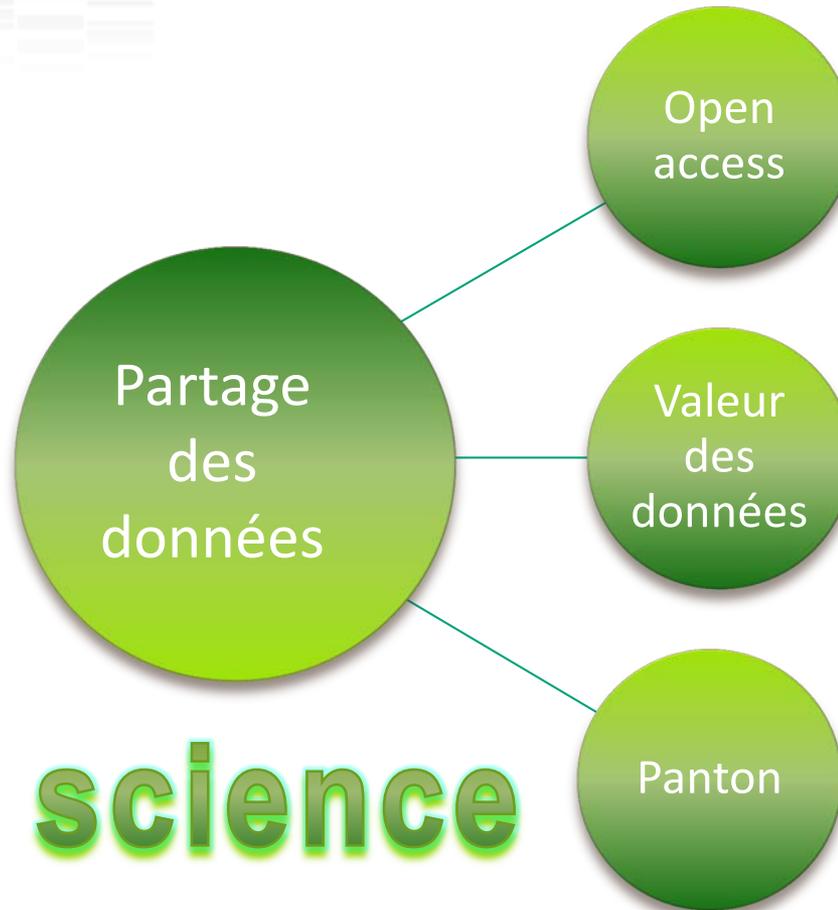
- ❖ 11 principes pour une **politique de partage des données de la recherche**
- ❖ Un projet inscrit au schéma directeur des systèmes d'information
 - ✓ Chantiers disciplinaires
 - ✓ Chantier juridique et déontologique
 - ✓ Chantier technique
 - ✓ Chantier compétences



_02

Contexte international

Les origines



- Déclaration de Berlin 2003 : Libre accès à l'IST publi/données

- OCDE 2007 : Recherche financée sur fonds publics

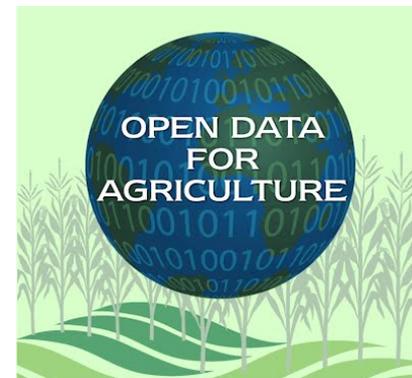
- Déclaration 2004 : Spécifique sur les données

Open science

Des catalyseurs récents

❖ G8+5 Open data for Agriculture

Obtain commitment and action from nations and relevant stakeholders to **promote policies and invest in projects that open access to publicly funded global agriculturally relevant data streams**, making such data readily accessible to users in Africa and world-wide, and ultimately supporting a sustainable increase in food security in developed and developing countries



❖ Research data Alliance

The Research Data Alliance aims to **accelerate and facilitate research data sharing and exchange**



Du côté de l'édition

❖ Des nouvelles revues :



BGI Shenzhen and
BioMed Central

GigaScience aims to revolutionize data dissemination, organization, understanding, and use. An online open-access open-data journal, **we publish 'big-data' studies from the entire spectrum of life and biomedical sciences.** To achieve our goals, the journal has a novel publication format: one that **links standard manuscript publication with an extensive database that hosts all associated data and provides data analysis tools and cloud-computing resources.**

Not just 'omic' type data ... imaging, neuroscience, ecology, cohort data, systems biology and other new types of large-scale sharable data.

❖ Note aux auteurs – revues classiques

The logo for the journal Science, consisting of the word 'Science' in white text on a red rectangular background.

Data and materials availability All data necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of *Science*.]...[*Science* supports the efforts of databases that aggregate published data for the use of the scientific community. Therefore, appropriate data sets (including microarray data, protein or DNA sequences, atomic coordinates or electron microscopy maps for macromolecular structures, and climate data) must be deposited in an approved database, and an accession number or a specific access address must be included in the published paper. We encourage compliance with MIBBI guidelines (Minimum Information for Biological and Biomedical Investigations).

❖ De nouveaux entrepôts de données



Dryad is an international repository of data underlying peer-reviewed articles in the basic and applied biosciences. Dryad enables scientists to validate published findings, explore new analysis methodologies, repurpose data for research questions unanticipated by the original authors, and perform synthetic studies. Dryad is governed by a consortium of journals that collaboratively promote data archiving and ensure the sustainability of the repository.

22 JUN 2012

THOMSON REUTERS UNVEILS DATA CITATION INDEX FOR DISCOVERING GLOBAL DATA SETS

First of Its Kind Data Citation Index Connects Researchers to Data Repositories around the World

Philadelphia, PA, June 22, 2012 — The Intellectual Property & Science division of Thomson Reuters announced today that it will preview at the American Library Association Conference (ALA) the *Data Citation Index™*, an upcoming research resource within the *Web of Knowledge™* to facilitate the discovery, use and attribution of data sets and data studies, and link those data to peer-reviewed literature.

This new research resource from Thomson Reuters creates a single source of discovery for scientific, social sciences and arts and humanities information by connecting foundational research within data repositories around the world to related peer-reviewed literature in journals, books, and conference proceedings already indexed in the *Web of Knowledge*.

The Thomson Reuters *Data Citation Index*, scheduled for release later this year, makes research within the digital universe discoverable, citable and seamlessly linked to the article detailing the outputs from the original investigation. Thomson Reuters has partnered with data repositories such as the Inter-University Consortium for Political and Social Research (ICPSR) to capture bibliographic records and cited references for digital research, facilitating visibility, author attribution, and ultimately the measurement of impact of this growing body of scholarship.

"We are excited to partner with Thomson Reuters in the building of the *Data Citation Index*," said Mary Vardigan, assistant director of the ICPSR. "By linking publications in the *Web of Science* to the datasets on which they are based and enhancing the discoverability of data through the *Data Citation Index*, Thomson Reuters is highlighting the importance of research data in the scientific process."

"The *Data Citation Index* will revolutionize the way data sets are discovered and utilized," said Keith MacGregor, executive vice president of Thomson Reuters. "It will

THE DATA CITATION INDEXSM

DEFINITIONS:

Data repository: a database or collection comprising data studies, data sets and/or microcitations which stores and provides access to the raw data. Constituent data studies, and sometimes individual data sets, are marked up with metadata providing a context for the available raw data.

Data study: description of studies or experiments held in repositories with the associated data which have been used in the data study. (Includes serial or longitudinal studies over time). Data studies can be a citable object in the literature and may have cited references attached in their metadata, together with information on such aspects as the principal investigators, funding information, subject terms, geographic coverage etc. The level of metadata provided varies between repositories.

Data set: a single or coherent set of data or a data file provided by the repository, as part of a collection, data study or experiment. Data sets may present in a number of file formats and media types: they may be number based files such as spreadsheets, images, video, audio, databases etc. Data sets can be a citable object in the literature and may have cited references attached in their metadata, but more commonly they inherit the metadata of the overall study in which they are used.



Autres faits marquants

❖ Révision du « Grant Proposal Guide » de la NSF : janvier 2013

Chapter II.C.2.f(i)(c), Biographical Sketch(es), has been revised to **rename the “Publications” section to “Products”** and amend terminology and instructions accordingly. This change makes clear that products may include, but are not limited to, publications, data sets, software, patents, and copyrights.

❖ Recommandations de la CE de juillet 2012

D’ici à 2014 des politiques de libre accès aux publications et aux données de la recherche auront été établies dans tous les états membres



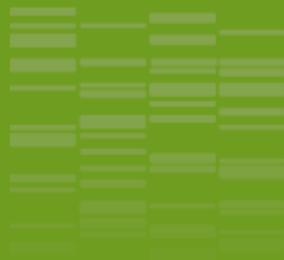
_03

Objectifs du séminaire

Votre mission ...

Objectifs

- ❖ Finaliser les principes
- ❖ Partager une culture de base commune
- ❖ Organiser le dispositif projet
 - ✓ participants
 - ✓ chantiers
 - ✓ cadre et démarche



_04

Programme du séminaire

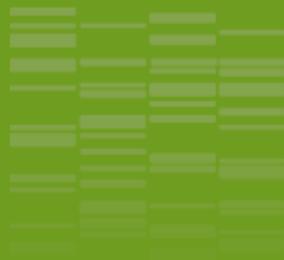
Conférences et ateliers

❖ Différents points de vue :

- ✓ Politique : Carlos Morais Pires, DG-Connect
- ✓ Chercheur : Salam Mesmoudi, Inserm
- ✓ Organisation inter-gouvernementale : Johannes Keizer, FAO
- ✓ Institution : James Wilson, Univ. Oxford
- ✓ Discipline : Françoise Génova, CDS Astronomie

❖ 3 types d'atelier

- ✓ Atelier 1 : Appropriation / échange autour des principes
- ✓ Atelier 2 : Questions à étudier / cahier des charges
- ✓ Atelier 3 : Organisation des chantiers



05

Partage de données

Forces –Faiblesses- Opportunités- Menaces

Forces

- ❖ Politique posée, implication DG, volontarisme des acteurs
- ❖ Patrimoine de données
- ❖ Existant de données structurées génomiques - données spatiales
- ❖ Données originales, pluridisciplinaires et systèmes intégrés
- ❖ Organisme vs agence
- ❖ Organisation interne : ORE, PF, UE, réseaux métiers, CATI
- ❖ Habitude de la collaboration
- ❖ Complémentarité des compétences scientifiques, IST, informatique
- ❖ Savoir faire en production de connaissances – linked data
- ❖ Capacité de leadership sur instances de normalisation/standards
- ❖ Démarche qualité en recherche
- ❖ Expérience des CRB sur conditions de mise à disposition
- ❖ Moyens sur les infrastructures

Faiblesses

- ❖ Ne pas être seulement un producteur de données
- ❖ Individualisme / réticence du chercheur, gestion données vue comme une contrainte
- ❖ Reconnaissance, évaluation
- ❖ Absence de définition de la donnée à partager
- ❖ Absence de gestion du cycle de vie de la donnée
- ❖ Faible capacité de gestion des données et faible niveau de structuration des données
- ❖ Qualité des données, qualité et description des protocoles
- ❖ Manque de standards, de normes, d'ontologie
- ❖ Ressources insuffisantes, moyens humains, compétences, coût
- ❖ Cadre juridique méconnu, mal maîtrisé notamment dans le cadre partenariats publics comme privés
- ❖ UMR et politiques différentes entre tutelles
- ❖ Compétition si les données sont partagées trop tôt

Opportunités

- ❖ Gagnant / gagnant, bénéficiaire d'un regard et d'un enrichissement extérieur
- ❖ Crowdsourcing, impliquant les agriculteurs
- ❖ Gagner en crédibilité, transparence des recherches, preuve
- ❖ Démonstration de l'utilité publique de la recherche
- ❖ Revaloriser des données anciennes
- ❖ Ouverture de nouveaux champs d'investigation
- ❖ Diminuer les coûts de recherche
- ❖ Réduire l'expérimentation animale (éthique)
- ❖ 'Pédagogie' du « pourquoi partager » vers chercheurs
- ❖ Renforcement démarche qualité de la recherche
- ❖ Être un acteur reconnu du partage de données
- ❖ Archivage, temps de conservation ?
- ❖ Développement des outils de méta-analyse, partager des outils
- ❖ Être capable d'intégrer des données de différentes sources et disciplines
- ❖ Être dans les instances de normalisation / standardisation
- ❖ Projets et financements européens, projets internationaux, ANR

Menaces

- ❖ Pillage - perte d'avantage compétitif avec les concurrents
- ❖ Mauvais usage par les médias, les citoyens ?
- ❖ Complexification juridique
- ❖ Injonction paradoxale : valorisation 'économique', brevets
- ❖ Perte de petits jeux de données
- ❖ Données sensibles
- ❖ Méta-analyse prématurée par d'autres
- ❖ Ne pas re-questionner politique/démarche régulièrement
- ❖ Coût de conservation
- ❖ Dimensionnement des équipements
- ❖ Évolution technologique

Pistes à explorer – questions émergentes (1/2)

- ❖ Pédagogie – communication sur les enjeux
- ❖ Données :
 - ✓ quelles données,
 - ✓ identifier les jeux de données, de quoi parle-t-on ?
 - ✓ identifier les jeux de données partageables
 - ✓ cycle de vie de la donnée
 - ✓ plan de gestion des données
 - ✓ qualité des données
 - ✓ archivage – temps de conservation
- ❖ Les standards existants, standards sur lesquels se positionner

Pistes à explorer – questions émergentes (2/2)

- ❖ Outil(s) de méta-analyse
- ❖ Plan de gestion compétences : structuration, description données, curation données, outils/infra...
- ❖ Plate-formes de dépôt, infrastructures existantes, à recommander, à créer
- ❖ Reconnaissance de la diffusion des données dans évaluation
- ❖ Propriété intellectuelle et cadre juridique contrat de recherche: mentions, droits, conditions de ré-utilisation
- ❖ Comment revisiter politique / démarche régulièrement
- ❖ Comité d'évaluation des données à partager (visibilité vs sensibilité...)

Quelle(s) approche(s)

- ❖ Par domaines de recherche, de données :
'omiques', agro-environnementales, SHS...
- ❖ Par métiers (transverses) : scientifique,
juridique, IST, informatique....



_06

Introduction des travaux ‘approche, démarche et cadre de travail’

Ateliers 2 & 3

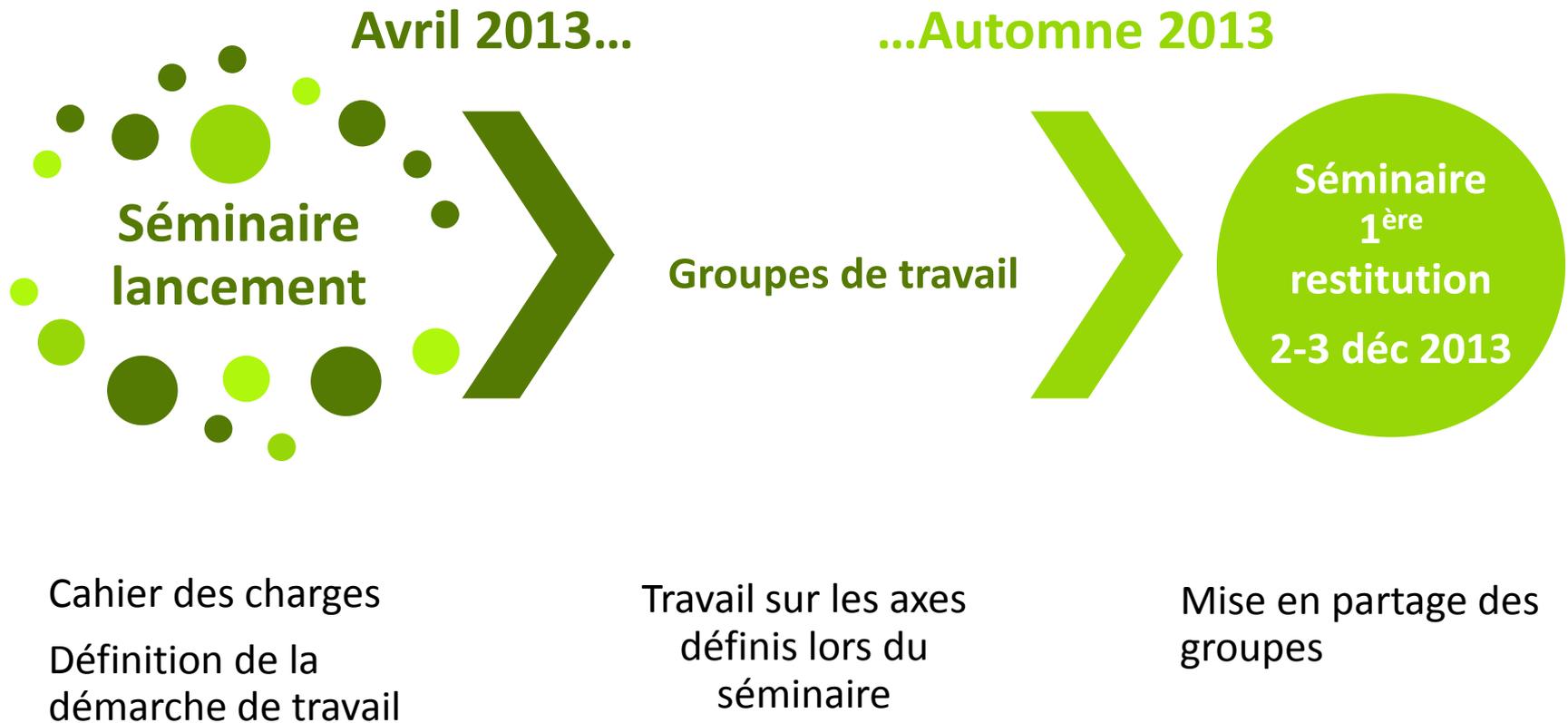
❖ Atelier 2 :

- ✓ Cahier des charges : questions à traiter pour définir le plan de mise en œuvre des principes
- ✓ Définition de l'approche : thématique ou disciplinaire - identification des groupes

❖ Atelier 3 :

- ✓ Cadrage des sujets à traiter
- ✓ Identification des livrables
- ✓ Modalités de travail

Planning et livrables





07

Démarche, cadre de travail et plan d'action à l'issue du séminaire

Élaboration de la stratégie de valorisation des données de la recherche

Objectifs de la démarche de travail

- ❖ Élaborer la stratégie de valorisation des données de recherche dans le cadre du partage
- ❖ Anticiper l'évolution des compétences pour accompagner la mise en œuvre du partage des données
- ❖ Définir les recommandations, méthodes, outils... pour la mise en œuvre du partage

1^{er} axe de travail 'données'

- ❖ Définition par le groupe pluridisciplinaire des questions à traiter (17 avril 2013)
- ❖ Travail sur les questions par groupe de famille de données selon nomenclature des SI stratégiques
<https://intranet6.inra.fr/systemes-information/Informatique-scientifique/Production>
groupes de travail mobilisant des scientifiques, informaticiens et documentalistes (restitution des groupes septembre 2013)
- ❖ Mise en commun des groupes de famille et travail en un groupe pluridisciplinaire pour définition de la stratégie de valorisation des données (point d'étape 1^{ere} restitution décembre 2013, puis mi 2014)

Axe données – questions à traiter (1/2)

dans un 1^{er} temps par 3 groupes de famille de données puis en pluridisciplinaire

1. Définir les cibles stratégiques

- ✓ Quels critères ? Analyse du contexte international (grand public, partenaires professionnels, scientifique) du domaine,
- ✓ Élément de preuve scientifique (grand public, communauté scientifique)
- ✓ Se doter d'une politique de gestion de données
- ✓ **Adaptation** de politique de gestion de données par famille de données

2. Quelles données partageons-nous?

- ✓ Que produit-on ? Avec qui ? Qu'est-ce qu'une donnée?
- ✓ Types de données et pour quels usages
- ✓ Identification des données en commençant par celles portées par actifs stratégiques de SI scientifiques
- ✓ Qualification de la donnée diffusable ? Brute/élaborée (Critères, modalités d'évaluation ?)
- ✓ Non numérisées, données historiques et patrimoniales

Axe données – questions à traiter (2/2)

dans un 1^{er} temps par 3 groupes de famille de données puis en pluridisciplinaire

3. Cycle de vie donnée – évolution de la donnée

- ✓ Production
- ✓ Partage
- ✓ Archivage
- ✓ (Destruction)
- ✓ Réincarnation?

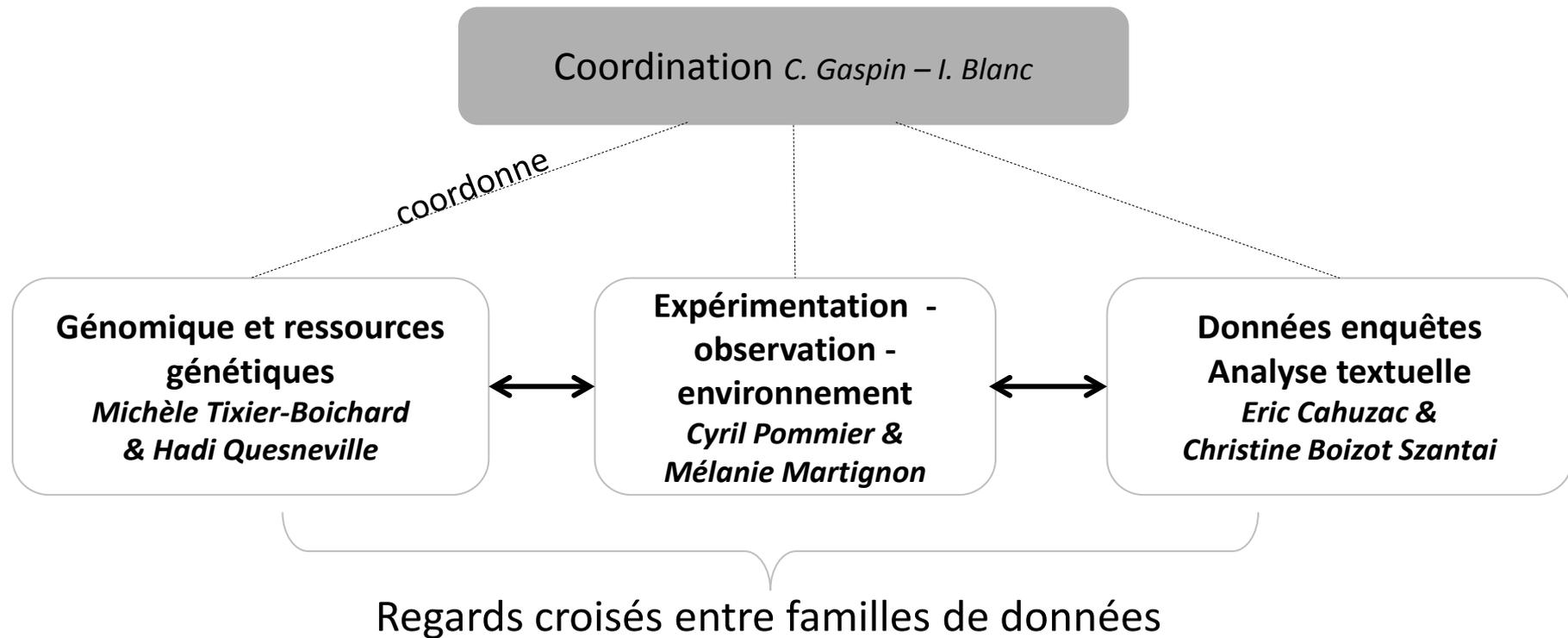
4. Comment diffuse-t-on ?

- ✓ Pratiques et standard selon familles de données
- ✓ Conditions de mise à disposition
- ✓ Modalités de dépôts
- ✓ BDD/PF spécifiques ou généralistes

Dispositif travail – Données

Travail à partir du cadre générique des questions

- ❖ approche par palier selon famille de données, contexte existant et urgence des questions



Génomique et ressources génétiques

Contributeurs

Michèle Tixier-Boichard & Hadi Quesneville, Sébastien Carrere, Jean-Louis Noyer (CIRAD), Manuel Ruiz (CIRAD), Christophe Klopp (?), Aurélie? ...

Périmètre

1. Définir les cibles stratégiques
2. Quelle données partageons-nous ?
3. Cycle de vie donnée – évolution de la donnée
4. Comment diffuse-t-on?

Principes de fonctionnement

1. État des lieux par famille (élargir les groupes *via* département)
2. Synthèse du travail par famille et mise en commun
3. Travail par question toutes familles confondues
4. Mise en commun des travaux
5. Partage du document avec les communautés

Livrables - Jalons

- Sept 2013 : Réunion de mise en commun entre familles
- Déc 2013 : Document « draft » en pluridisciplinaire à partager avec méthodes et outils
- 2014 : Document final

Expérimentation - observation - environnement

Contributeurs

Cyril Pommier & Mélanie Martignon, Jean-François Hoquette, Jérôme Molénat, Denis Loustau, Marie-Colette Fauré, Virginie Lelièvre, Daniel, Jacob, François Laperruque, Marc Deconchat, Christian Ducrot, Isabelle Lebret, Jean-François Rami (CIRAD) ...

Périmètre

1. Définir les cibles stratégiques
2. Quelle données partageons-nous ?
3. Cycle de vie donnée – évolution de la donnée
4. Comment diffuse-t-on?

Principes de fonctionnement

1. État des lieux par famille (élargir les groupes *via* département)
2. Synthèse du travail par famille et mise en commun
3. Travail par question toutes familles confondues
4. Mise en commun des travaux
5. Partage du document avec les communautés

Livrables - Jalons

- Sept 2013 : Réunion de mise en commun entre familles
- Déc 2013 : Document « draft » en pluridisciplinaire à partager avec méthodes et outils
- 2014 : Document final

Données enquêtes - Analyse textuelle

Contributeurs

Eric Cahuzac & Christine Boizot Szantai, Christophe Bontemps, Véronique Batifol-Garandel, Pierre Triboulet , Patrice Buche, Philippe, Breucker, Véronique Coxam, Lydie Soler

Périmètre

1. Définir les cibles stratégiques
2. Quelle données partageons-nous ?
3. Cycle de vie donnée – évolution de la donnée
4. Comment diffuse-t-on?

Principes de fonctionnement

1. État des lieux par famille (élargir les groupes *via* département)
2. Synthèse du travail par famille et mise en commun
3. Travail par question toutes familles confondues
4. Mise en commun des travaux
5. Partage du document avec les communautés

Livrables - Jalons

- Sept 2013 : Réunion de mise en commun entre familles
- Déc 2013 : Document « draft » en pluridisciplinaire à partager avec méthodes et outils
- 2014 : Document final

2nd axe de travail 'Méthodes et outils'

Coordination : O. Hologne – P. Neveu

Mission

Structuration et interopérabilité – Pascal Neveu

Formats de données, Modèles de données, Métadonnées, Ontologies, standards, identifiants, sécurité...

Mission

Méthodes pour l'exploitation des données – Alain Trubuil

Méta-analyse, analyse statistique, data-mining, inférence, raisonnement, évaluation, visualisation

Structuration et interopérabilité

Contributeurs

Pascal Neveu, Sophie Aubin, Pascal Aventurier, Juliette Dibie-Barthélémy, David Crevoisier (?), Nicolas Saby (?), Marc Irvine(?), équipe Wageningen, Esther Dzalé-Yeumo – Sébastien Carrere, Jean-Yves Lorendeau, Inria- équipe Zenith

Périmètre

Groupe orienté 'producteurs de données'

- Métadonnées : provenance, recommandations
- Formats, modèles de données
- Identifiants
- Ontologies
- Mise à disposition, dissémination
- Découverte des jeux de données
- Portail, annuaire
- Sécurisation

Principes de fonctionnement

- Travail en groupe
- Croisement avec les 3 familles de données

Livrables - Jalons

- Octobre 2013 : état des lieux
- Février 2014 : recommandations et standards sur la structuration des données
- Été 2014 : Guide décliné pour les 3 familles de données

Méthodes pour l'exploitation des données

Contributeurs

Alain Trubuil, Patrice Buche, Pascal Aventurier, Claire Chabanet, Christelle Hennequet, Patrick Gasqui, David Legland, Christophe Bontemps, Data publica

Périmètre

Groupe orienté 'valorisation – publication des données'

- Méta-analyse
- Analyse statistique
- Data-mining
- Inférence
- Raisonnement
- Évaluation
- Visualisation

Principes de fonctionnement

- Travail en groupe
- Février 2014 : 1^{ère} version recommandations à croiser avec groupes données

Livrables - Jalons

- Octobre 2013 : état des lieux des méthodes d'exploitation
- Été 2014 : Recommandations
 - Quelles méthodes d'exploitation pour les données ouvertes ?
 - Comment utiliser et partager méthodes ?

Chantiers transverses

- Stratégie : définition de la stratégie de valorisation des données de recherche à partir des synthèses des travaux/recommandations des groupes données et méthodes-outils
- Communication – sensibilisation des scientifiques
 1. Sur les enjeux et principes en matière de partage de la donnée
 2. La démarche d'élaboration de la stratégie de valorisation
 3. La stratégie de la valorisation
 4. Les moyens pour accompagner cette stratégie
- PI et aspects juridiques Conditions de mise à disposition, licences, creative commons, des contrats / conventions...
- Reconnaissance des acteurs (chercheurs et opérateurs du partage...)
- Compétences, formations, et accompagnements
- Data citation index, données liées aux publications, stratégies éditeurs, place de Prodnra 2.0

Stratégie et communication

Contributeurs

Christian Pichot, Erwin Dreyer, Pascal Neveu, Odile Hologne, Christine Gaspin, Isabelle Blanc

Périmètre

- **Communication – sensibilisation**
 - Sur les enjeux et principes en matière de partage de la donnée
 - La démarche d'élaboration de la stratégie de valorisation
 - La stratégie de la valorisation
 - Les moyens pour accompagner cette stratégie
- **Synthèse des travaux des axes données et outils méthodes :**
 - État de l'art et positionnement (PF, choix politiques...)
 - Élaboration de la stratégie de mise en œuvre : plan de gestion des données
- **Reconnaissance des acteurs engagés dans partage**

Principes de fonctionnement

- **Communication sur les principes et politique de mise en œuvre**
 - Communication enjeux-principes (5 juillet 2013)
 - Communication sur avancées des travaux
- **Synthèse des travaux des 2 axes, pour élaborer stratégie Inra**

Livrables - Jalons

- Juillet 2013 : communication des enjeux et des principes et sur le plan d'action de l'élaboration de la stratégie
- Fin 2013 : Comm sur les travaux en cours 1ere restitution états des lieux – état de l'art
- Été 2014 : Communication sur stratégie de mise en œuvre
- Au-delà : communication sur les moyens mis en œuvre pour accompagner le partage : formation, guides de recommandations, infrastructures...

PI et aspects juridiques

Contributeurs

En cours d'identification

Périmètre

Groupe inter-organismes

- Les données dans les contrats
- Licences de diffusion des données
- ...

Principes de fonctionnement

Livrables - Jalons

État de l'art, étude d'opportunité : Data citation index, données liées aux publications, stratégies éditeurs, place de Prodinra...

Contributeurs

Esther Dzale, Dominique L'Hostis, Pascal Aventurier, Véronique Garandel, Diane Le Hénaff, Marie-Colette Fauré, Virginie Lelièvre, ...

Périmètre

- Data journals : type d'articles, reviewing, stockage des données
- Mise en œuvre des DOI
- Dépôt centralisé des données publiées : Prodinra2.0 ou autre alternative
- Citation des données : conseils aux auteurs
- Data citation index : Compréhension du fonctionnement, conseils à donner
- Repérage des bonnes ou mauvaises pratiques des éditeurs
- Évaluation des nouveaux entrepôts de données externes (Dryad, Pangaea)

Principes de fonctionnement

Étude pour produire analyse des points du périmètre voire étude d'opportunité

Binômes à identifier en charge de chaque item
Échanges avec les axes données et méthodes-outils

Livrables - Jalons

- Chaque item fait l'objet d'un livrable. L'ensemble sera présenté au séminaire du 2 et 3 décembre 2013

Compétences – formations et accompagnement

Contributeurs

Laurent Bruckler, xx, FPN

Périmètre

- Identification des compétences nouvelles
- Plan de gestion prévisionnelle de ces compétences
- Formation 'écoles thématiques open data' jeunes chercheurs, DU, documentalistes, informaticiens
- Élaboration d'un plan et d'une organisation d'accompagnement des chercheurs

Principes de fonctionnement

- Démarrage après 1^{er} séminaire de restitution – décembre 2013

Livrables - Jalons

Questions ouvertes

Questions remontées lors des ateliers et qui seront traitées après avoir mieux cerné les données à partager, les conditions de mise à disposition, l'utilisation d'entrepôt existant... et commencer à élaborer stratégie :

- ✓ Stockage
- ✓ Conservation
- ✓ P2P data sharing
- ✓ Plan de gestion des données

Démarrage 1^{er} semestre 2014 sur la base des pré-recommandations de la stratégie de valorisation des données

Conclusions Olivier Le Gall (1/2)

- ❖ La communication sur la dynamique du partage doit être portée à l'échelle de l'Institut
- ❖ Importance de la présence du Cirad aux côtés de l'Inra dans cette démarche
- ❖ Dynamique doit interpeller les producteurs comme ceux qui analysent les données
- ❖ Reconnaissance individuelle dans le cadre du partage doit donc toucher autant les producteurs que ceux qui valorisent les données en publiant
- ❖ Des plateformes type Genbank dans des domaines comme l'environnement sont à développer
- ❖ Le métier de documentaliste doit poursuivre sa mutation

Conclusions Olivier Le Gall (2/2)

- ❖ Cette dynamique d'ouverture et de partage va transformer la culture de l'établissement
- ❖ « Cette dynamique va impacter de manière importante et durable l'image que l'on a de soi en tant que chercheur et comme chercheur dans la société »
- ❖ **Ce séminaire ouvre un chantier fondateur** pour le partage des données de la recherche de l'Inra

Prochaines étapes

- ❖ Finaliser la constitution des groupes de travail
- ❖ Préciser les questions à traiter pour les 2 missions sur axe Méthodes et outils
- ❖ 19 juin 2013 : point de coordination des 3 groupes 'données' et 2 groupes 'outils et méthodes' pour produire le plan d'action définitif
- ❖ 25 septembre 2013 : point de coordination des 3 groupes 'données' et 2 groupes 'outils et méthodes' – point d'avancement
- ❖ 2 & 3 décembre 2013 : séminaire réunissant la totalité des groupes pour une 1^{ère} restitution et calage des travaux à conduire pour l'élaboration de la stratégie et des guides de recommandations
 - ✓ Lancement du groupe 'compétences – formation – accompagnement'
- ❖ Mi 2014 : diffusion de la stratégie de valorisation des données au travers de leur partage

Fonctionnement dispositif

Légende

Membres fixes

Membres variables en fonction des projets

MEMBRES



RÔLE



Planning



Coordination

- animateurs axes
- animateurs axe données
- animateurs axe méthodes et outils

- Coordination et regards croisés sur les questions des familles de données et méthodes et outils

- Juin 2013 : validation plan de travail et groupes
- Sept 2013 : point d'étape
- Déc 2013 : Croisement des travaux – 1ere restitution
- Mars 2014 : point d'étape reco

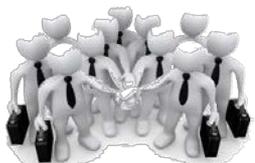


Axe données

- animateurs
- contributeurs
- **Invités opérationnels use case**

- Traite questions
- Rédige draft et
-

- Septembre 2013 :
- croisement des réflexions, résultats des 3 familles
 - définition des questions à traiter toutes familles confondues
- Déc 2013 : 1^{er} draft pour mise en partage
Mi 2014 recommandations



Axe outils et méthodes

- animateurs
- contributeurs
- Partenariats externes

- Traite questions
-

- Octobre 2013 :
- croisement état des lieux des 3 groupes
- Février 2014 :
- 1^{er} draft recommandations
- Mi 2014 recommandations

Planning stratégie et communication

Communication à chaque jalon

- ✓ Sortie séminaire + plan d'action définitif (juin 2013)
 - Communication enjeux et principes
 - Communication périmètres travaux et planning d'élaboration stratégie
- ✓ Décembre 2013 après séminaire 1^{ère} restitution
 - État de l'art par familles de données et approche pour traitement des questions en interdisciplinaire
 - État des lieux sur structuration des données, méthodes d'analyse....
 - Recommandations/ guide des études d'analyse ou d'opportunité

Élaboration de la stratégie de manière itérative au fil des avancées des travaux

Définition stratégie valorisation données – Planning prévisionnel

